form a regular lattice in the latent space. The centers $\boldsymbol{\zeta}_i$ of the delta functions are called the *latent vectors* of the GTM, and they are the GTM equivalent to the SOM map units. The approximation of the density generated in the data space is thus given by

$$p(\boldsymbol{\xi}|\mathbf{M},\beta) = \frac{1}{K}\sum_{i=1}^{K} p(\boldsymbol{\xi}|\boldsymbol{\zeta}_i;\mathbf{M},\beta) \tag{5}$$

The parameters of the GTM are determined by minimizing the negative log likelihood error

$$\mathcal{E}(\mathbf{M},\beta) = -\sum_{t=1}^{T} \ln\left[\frac{1}{K}\sum_{i=1}^{K} p(\boldsymbol{\xi}^t|\boldsymbol{\zeta}_i;\mathbf{M},\beta)\right] \tag{6}$$

over the set of sample vectors $\{\boldsymbol{\xi}^t\}$. The batch version of the GTM uses the EM algorithm [4]; for details, see [3]. One may also resort to an on-line gradient descent procedure that yields the GTM update steps

$$\boldsymbol{\mu}_j^{t+1} := \boldsymbol{\mu}_j^t + \delta^t\beta^t \sum_{i=1}^{K} \eta_i^t(\mathbf{M}^t,\beta^t)\phi_j(\boldsymbol{\zeta}_i)[\boldsymbol{\xi}^t - \boldsymbol{v}(\boldsymbol{\zeta}_i;\mathbf{M}^t)] \tag{7}$$

$$\beta^{t+1} := \beta^t + \delta^t\left[\frac{1}{2}\sum_{i=1}^{K} \eta_i^t(\mathbf{M}^t,\beta^t)||\boldsymbol{\xi}^t - \boldsymbol{v}(\boldsymbol{\zeta}_i;\mathbf{M}^t)||^2 - \frac{D}{2\beta^t}\right] \tag{8}$$

where $\eta_i^t(\mathbf{M},\beta)$ is the GTM counterpart to the SOM unit activation, the posterior probability $p(\boldsymbol{\zeta}_i|\boldsymbol{\xi}^t;\mathbf{M},\beta)$ of the latent vector $\boldsymbol{\zeta}_i$ given data vector $\boldsymbol{\xi}^t$:

$$\begin{aligned}\eta_i^t(\mathbf{M},\beta) &= p(\boldsymbol{\zeta}_i|\boldsymbol{\xi}^t;\mathbf{M},\beta)\\ &= \frac{p(\boldsymbol{\xi}^t|\boldsymbol{\zeta}_i;\mathbf{M},\beta)}{\sum_{i'=1}^{K} p(\boldsymbol{\xi}^t|\boldsymbol{\zeta}_{i'};\mathbf{M},\beta)}\\ &= \frac{\exp[-\frac{\beta}{2}\|\boldsymbol{v}(\boldsymbol{\zeta}_i;\mathbf{M}) - \boldsymbol{\xi}^t\|^2]}{\sum_{i'=1}^{K}\exp[-\frac{\beta}{2}\|\boldsymbol{v}(\boldsymbol{\zeta}_{i'};\mathbf{M}) - \boldsymbol{\xi}^t\|^2]}\end{aligned} \tag{9}$$

2.3 Connections between SOM and GTM

Let us consider a GTM that has an equal number of latent vectors and basis functions[1], each latent vector $\boldsymbol{\zeta}_i$ being the center for one Gaussian basis function $\phi_i(\boldsymbol{\zeta})$. Latent vector locations may be viewed as units of the SOM, and consequently the basis functions may be interpreted as connection strengths between the units. Let us use the shorthand notation $\phi_j^i \equiv \phi_j(\boldsymbol{\zeta}_i)$. Note that $\phi_j^i = \phi_i^j$, and assume that the basis functions be normalized so that $\sum_{j=1}^K \phi_j^i = \sum_{i=1}^K \phi_j^i = 1$.

At the zero-noise limit, or when $\beta \to \infty$, the softmax activations of the GTM given in (9) approach the winner-take-all function (1) of the SOM. The winner unit $\boldsymbol{\zeta}_{c(t)}$ for the data vector $\boldsymbol{\xi}^t$ is the map unit that has its *image* closest to the data vector, so that the index $c(t)$ is given by

$$c(t) = \operatorname*{argmin}_i \left\|\boldsymbol{v}(\boldsymbol{\zeta}_i) - \boldsymbol{\xi}^t\right\| = \operatorname*{argmin}_i \left\|\left(\sum_{j=1}^{K}\phi_j^i\boldsymbol{\mu}_j\right) - \boldsymbol{\xi}^t\right\| \tag{10}$$

[1]Note that this choice serves the purpose of illustration only; to use GTM properly, one should choose much more latent vectors than basis functions.